Short Communication

Data-Driven Innovation in Workforce Selection: A Clustering-Based Workflow for Technology Adoption in Indonesian Construction SMEs

Daniel Jesayanto Jaya[1,2], Haryanto Haryanto[3], Putu Sudira[1], Nuryadin Eko Raharjo[4]

1. Technology and Vocational Education and Training Department, Universitas Negeri Yogyakarta, Indonesia; 2. Building Engineering Education Department, Universitas Negeri Jakarta, Indonesia; 3. Electrical Engineering Education Department, Universitas Negeri Yogyakarta, Indonesia; 4. Civil Engineering and Planning Education Department, Universitas Negeri Yogyakarta, Indonesia

Recruitment represents a critical socio-economic process that shapes labor market outcomes and organizational performance, especially for small and medium-sized enterprises (SMEs) operating in the construction sector. This study demonstrates how a data-driven and computational approach—using K-Means clustering—can support transparent, efficient, and innovation-oriented recruitment. A dataset of 30 shortlisted applicants from Indonesian construction consulting SMEs was evaluated across three competencies: AutoCAD drafting, report writing, and adaptability. The clustering process classified candidates into *Rejected*, *Under Consideration*, and *Accepted* groups, with validity supported by a Davies–Bouldin Index of 0.41. Unlike threshold-based evaluations, clustering reveals nuanced groupings across technical and soft skills, enabling more precise and fair workforce selection. Conceptually, this work contributes to innovation management in SMEs by illustrating how data-driven decision tools can enhance human capital selection, reduce bias, and promote technology adoption for better workforce planning.

Corresponding author: Daniel Jesayanto Jaya, danieljesayanto.2023@student.uny.ac.id

Background and Motivation

The construction industry in Indonesia—dominated by small and medium-sized consulting firms—plays a crucial role in economic and infrastructural development. Projects such as the Nusantara Capital City

 doi.org/10.32388/TGR1NV.2

(IKN) are expected to require more than 260,000 workers by 2024[1]. Recruitment in this context is not only a managerial and logistical challenge but also a social and economic process that shapes access, equity, and employability outcomes. Traditional selection practices are often constrained by subjective judgments, limited transparency, and potential bias, which can result in mismatched skills and lower productivity[2][3][4].

Computational methods offer opportunities to address these limitations by systematically analyzing candidate competencies and uncovering patterns invisible to conventional evaluation. Unlike threshold-based assessments, clustering identifies nuanced groupings across technical and behavioral attributes, providing a more equitable and efficient foundation for workforce selection[5].

Beyond their technical contribution, such approaches demonstrate how innovation management practices and data analytics can optimize recruitment efficiency, enhance human capital utilization, and support technology adoption in SMEs. Integrating these tools into daily management practices reflects a growing trend of digital transformation in small enterprises seeking agility and competitiveness in rapidly evolving markets[6][7].

This study contributes to this emerging discourse by presenting a reproducible clustering-based workflow applied to recruitment in Indonesian construction SMEs. The objective is not to propose new clustering theory but to demonstrate practical, innovation-oriented decision-support methods that can reduce bias and enhance fairness in workforce analytics.

Data Description

1. Sample

The dataset consists of 30 shortlisted applicants drawn from 161 total candidates applying to an Indonesian construction consulting SME. The sample represents the final stage of the recruitment process, where decisions carry significant consequences for both organizational performance and individual career opportunities. Although small, this dataset illustrates how computational approaches can increase transparency and equity in real-world workforce selection.

2. Competency Dimensions

Each applicant was assessed across three core competencies combining technical and behavioral skills crucial for employability in construction projects:

a. *AutoCAD drafting* – technical design and planning skill.

b. *Planning and supervision report writing* – communication and documentation skill.

c. *Adaptability* – soft skill representing teamwork and responsiveness to field challenges.

3. Data format

Scores were recorded on a 0–100 scale and standardized to ensure comparability. This structure allows transparent replication and analysis across candidates.

4. Clustered groups (K=3)

Using K-Means clustering, applicants were grouped into three categories:

a. **Cluster 1 – Rejected**: candidates scoring below technical and adaptability thresholds.

b. **Cluster 2 – Under Consideration**: candidates with moderate performance and clear potential for further training.

c. **Cluster 3 – Accepted**: candidates with consistently high scores across both technical and soft skills.

Cluster	Label	Mean AutoCAD Drafting	Mean Report Writing	Mean Adaptability
1	Rejected	68.1	72.0	71.2
2	Under Consideration	74.5	80.3	83.1
3	Accepted	90.2	83.6	89.4

Table 1. *Cluster Characteristics of Workforce Selection Outcomes*

These clusters highlight competency-based distinctions among applicants, illustrating structured workforce selection outcomes in construction SMEs.

Methods and Technical Validation

1. Preprocessing

Incomplete records were removed, and all scores normalized to 0–1. Normalization ensures fairness and minimizes bias from varying scoring ranges—treating recruitment as a structured, data-driven process.

2. Clustering

K-Means clustering was executed in R (using *stats*, *ggplot2*, and *plotly* packages). The optimal cluster number (k = 3) was determined via the elbow method. Its efficiency and interpretability make K-Means ideal for SMEs adopting lightweight analytics systems[5].

Comparative Note on Clustering Methods

To strengthen the methodological justification, several alternative clustering techniques were considered. Hierarchical clustering offers intuitive dendrogram-based interpretability but is less suitable for rapid iteration and tends to be computationally heavier for incremental SME-level data. Density-based methods such as DBSCAN are effective in identifying noise and arbitrary cluster shapes; however, they require parameter tuning that may be challenging for non-expert users and can be unstable with small datasets. Gaussian Mixture Models (GMM) provide probabilistic flexibility but rely on distributional assumptions uncommon in competency-driven recruitment data. In contrast, K-Means delivers high interpretability, low computational cost, and ease of replication, making it an ideal choice for SMEs adopting lightweight analytics tools for workforce selection.

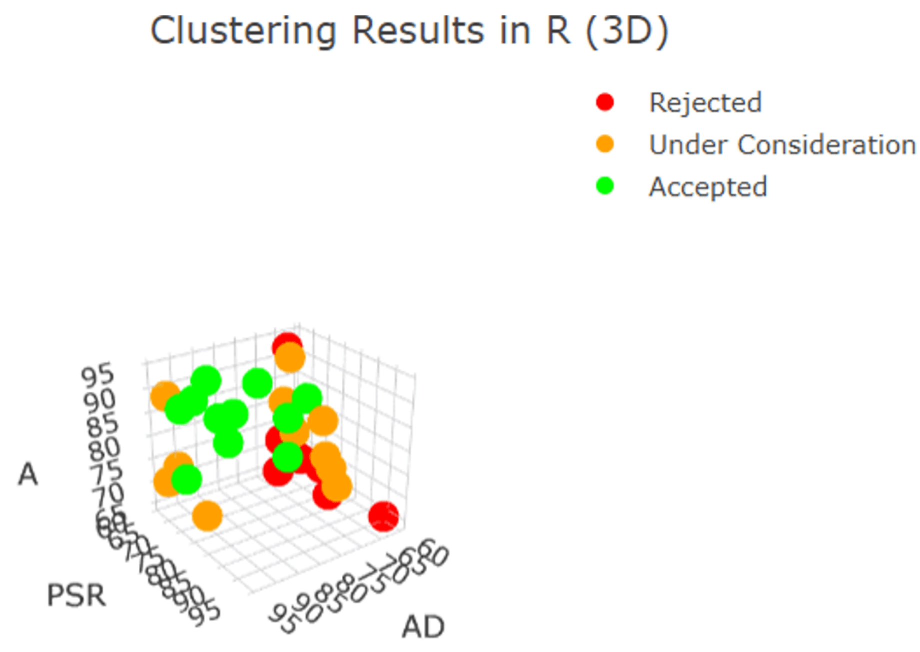
3. Validation

Cluster validity was measured using the Davies–Bouldin Index (DBI = 0.41), indicating well-separated and consistent group structures. This validates that the clusters represent meaningful competency distinctions relevant to managerial decision-making.

4. Visualization

Figures 1. *Two-dimensional scatter plot showing clusters of applicants by AutoCAD drafting and report-writing competencies.*

The plot visualizes trade-offs between technical competencies influencing recruitment outcomes.



Figures 2. *Three-dimensional scatter plot of all competency dimensions.*

The 3D visualization highlights how clustering reveals hidden applicant structures, increasing transparency in decision-making.

Usage Notes

K-Means clustering provides SMEs with a low-cost, replicable, and innovation-oriented analytics tool for understanding workforce dynamics. Compared to threshold-based evaluation, it captures nuanced patterns across technical and soft skills—helping employers identify balanced candidates and reduce subjectivity[5].

From an innovation management perspective, this workflow demonstrates how computational methods can enhance managerial decision-making, workforce allocation, and human capital development. For SMEs adopting digital tools, such methods represent a tangible step toward data-driven management and sustainable innovation practices[2][3].

doi.org/10.32388/TGR1NV.2

Practical Implications of Cluster Profiles

The three cluster groups offer actionable insights for construction SMEs seeking to optimize workforce allocation and training investments. Candidates in **Cluster 1 (Rejected)** demonstrate foundational gaps in both technical and adaptability competencies. For these applicants, organizations may consider targeted entry-level interventions, such as introductory AutoCAD training, basic documentation workshops, or short soft-skill modules, should they wish to build a pre-employment development pipeline.

Cluster 2 (Under Consideration) represents individuals with moderate proficiency and strong potential for upskilling. This group is well suited for structured microlearning programs, supervised internship placements, or project-based mentoring designed to accelerate competency convergence toward industry expectations. Focused interventions in adaptability and task documentation can yield rapid performance gains.

Finally, **Cluster 3 (Accepted)** consists of high-performing candidates who exhibit consistent strength across all competency dimensions. Organizations can strategically allocate these individuals to roles requiring advanced drafting, quality reporting, or field coordination responsibilities. Clear onboarding pathways, retention strategies, and capability-matching can further enhance their contribution to technology-driven project execution.

Broader implications include:

1. Fair recruitment – identifying balanced profiles reduces bias and improves organizational credibility.
2. Workforce development – medium-performing candidates (*Cluster 2*) can be trained strategically to address skill gaps.
3. Policy linkage – supports vocational–industry alignment by quantifying real-world competency demand.
4. Innovation adoption – builds managerial capacity to integrate analytics in everyday SME operations.

Limitations

The dataset is small (30 applicants, one SME), which limits generalizability. However, it demonstrates transparency and reproducibility as key innovation management principles for resource-constrained enterprises. Future research could scale up the dataset, incorporate additional indicators (e.g., leadership, digital skills), and test advanced clustering or AI-driven approaches[5][6][7].

Ethical Considerations and Algorithmic Bias

While clustering promotes transparency in workforce selection, algorithmic bias remains a potential risk. The quality and representativeness of the underlying dataset strongly influence cluster outcomes; any historical imbalances or structural inequities embedded in the data may be reproduced by the model. Although normalization helps mitigate technical bias, it does not fully eliminate socio-demographic biases. Therefore, clustered outputs should be complemented by human judgement, periodic audits, and fairness checks to ensure that data-driven recruitment supports equitable decision-making.

Closing Note

This study showcases how SMEs can integrate data analytics into human resource management as part of their innovation journey. The proposed workflow is transparent, reproducible, and economically feasible, offering a framework for technology adoption and human capital optimization.

Enhanced Theoretical and Practical Contribution

This study extends the discourse on innovation management in SMEs by demonstrating how computational workforce analytics can strengthen strategic HR decision-making. Beyond optimizing recruitment efficiency, the clustering workflow illustrates how data-driven tools can contribute to the development of organizational innovation capability—particularly in resource-constrained environments where technology adoption barriers remain significant. By integrating low-cost analytics into routine managerial practices, SMEs can gradually transition toward evidence-based workforce planning and strengthen their readiness for digital transformation.

Overall, this work contributes to the ongoing dialogue on innovation management in SMEs by illustrating how data-driven decision systems can improve workforce efficiency and support sustainable technological transformation[2][3].

About the Authors

Daniel Jesayanto Jaya is a Ph.D. candidate at the Graduate School, Universitas Negeri Yogyakarta (UNY), and a lecturer at the Building Engineering Education Study Program, Faculty of Engineering, Universitas

Negeri Jakarta (UNJ), Indonesia. His research interests focus on vocational education, career development, and the integration of intelligent decision-support systems in workforce analytics—particularly within the construction sector. He is the corresponding author of this article and can be contacted at danieljesayanto.2023@student.uny.ac.id. His ORCID is 0000-0003-1940-6302.

Dr. Haryanto is an associate professor at the Faculty of Engineering, Universitas Negeri Yogyakarta (UNY), Indonesia. His research interests include artificial intelligence control, educational research and evaluation, and technical and vocational education and training (TVET). He can be contacted at haryanto@uny.ac.id, and his Scopus ID is 57217152392.

Prof. Dr. Putu Sudira is a professor specializing in vocational pedagogy and industry–education collaboration. He can be contacted at putupanji@uny.ac.id, and his ORCID is 0000-0003-2646-6741.

Dr. Nuryadin Eko Raharjo is an associate professor in civil engineering education focusing on sustainable construction and modern training technologies. He can be contacted at nuryadin_er@uny.ac.id, and his ORCID is 0009-0002-9847-8987.

Statements and Declarations

Conflicts of Interest

The authors declare no competing interests.

Funding

This research was supported by the Indonesia Endowment Fund for Education (LPDP), Grant Number 202312211239164.

Ethical Approval

The study used anonymized recruitment data provided under a formal agreement. No human participants were directly involved, and ethical clearance was not required under institutional guidelines.

Informed Consent

Permission to use the anonymized dataset was obtained through a signed agreement with the data provider.

Author Contributions (CRediT taxonomy)

Daniel Jesayanto Jaya: Conceptualization, Methodology, Data Curation, Formal Analysis, Writing – Original Draft.

Putu Sudira: Supervision, Writing – Review & Editing.

Nuryadin Eko Raharjo: Methodology, Validation, Writing – Review & Editing.

Data Availability

The dataset and R workflow used in this study (including preprocessing, clustering, and validation scripts) are openly available on Zenodo: https://doi.org/10.5281/zenodo.16936952.

A synthetic dataset mirroring the original structure is provided to ensure confidentiality while allowing full reproducibility.

Acknowledgements

The authors gratefully acknowledge the support of the Indonesia Endowment Fund for Education (LPDP), which provided doctoral scholarship funding that enabled this research and manuscript preparation.

References

1. ^*Supriyanti SS, Kusmayanti JD, Paluseri ARA, Syahid CN, Oktarina A, Aziz NLL (2023). "Pemberdayaan Masyarakat Sekitar di Wilayah Ibu Kota Nusantara" [Community Empowerment in the Nusantara Capital City Region]. Masyarakat Indonesia.* ***49****(1):93–102. https://ejournal.brin.go.id/jmi/article/view/8528.*
2. a, b, c*Akkermans J, Donald WE, Jackson D, Forrier A (2024). "Are We Talking About the Same Thing? The Case for Stronger Connections Between Graduate and Worker Employability Research." CDI.* ***29****(1):80–92. doi:10.1108/CDI-08-2023-0278.*
3. a, b, c*Donald WE, Van der Heijden BIJM, Manville G (2024). "(Re)Framing Sustainable Careers: Toward a Conceptual Model and Future Research Agenda." CDI.* ***29****(5):513–526. doi:10.1108/CDI-02-2024-0073.*
4. ^*Brown P, Hesketh A (2005). The Mismanagement of Talent: Employability and Jobs in the Knowledge Economy. Ind Labor Relat Rev. doi:10.2189/asqu.2005.50.2.306.*
5. a, b, c, d*Jain AK, Murty MN, Flynn PJ (1999). "Data Clustering: A Review." ACM Comput Surv.* ***31****(3):264–323. doi:10.1145/331499.331504.*

6. [a, b]*Lazer D, Pentland A, Adamic L, Aral S, Barabási A-L, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, Jebara T, King G, Macy M, Roy D, Van Alstyne M (2009). "Computational Social Science." Science.* ***323****(5915):721–723. doi:10.1126/science.1167742.*
7. [a, b]*Conte R, Gilbert N, Bonelli G, Cioffi-Revilla C, Deffuant G, Kertesz J, Loreto V, Moat S, Nadal J-P, Sanchez A, Nowak A, Flache A, San Miguel M, Helbing D (2012). "Manifesto of Computational Social Science." Eur Phys J Spec Top.* ***214****(1):325–346. doi:10.1140/epjst/e2012-01697-8.*

Supplementary data: available at https://doi.org/10.32388/TGR1NV.2

Declarations

Funding: This research was supported by the Indonesia Endowment Fund for Education (LPDP), Grant Number 202312211239164.

Potential competing interests: No potential competing interests to declare.